

# Intelligent Archive

*Integrated Data Access and Organization for Scientists*

## Technology

The Intelligent Archive (IA) provides scientists with advanced capabilities for organizing, searching, and interacting with information and data. The IA environment integrates custom LLNL-developed software with commercial and public-domain software such as database systems and World Wide Web browsers.

## Applications

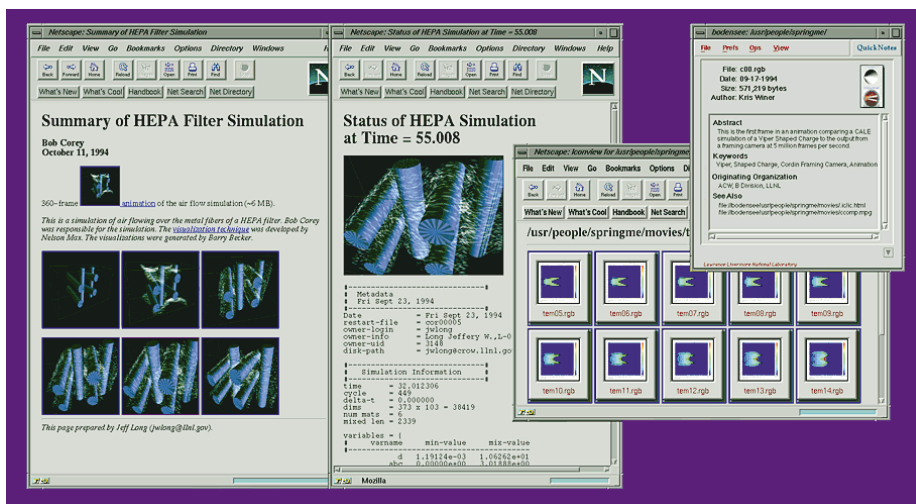
IA tools provide data access and organization for:

- Diverse types of data.
- Distributed resources.
- Shared information.

Scientists require access to disparate types of information, including computed and experimental data, papers, reports, and notes. The IA project began in January 1993 in response to this requirement, investigating ways to provide intelligent access to archival storage systems, hence the "Intelligent Archive" name. The scope of the IA has expanded to access data in more locations, with more capabilities, for more customers. Our current goal is to advance scientists' capabilities for organizing, searching, and interacting with information and data.

## The IA Approach

The IA allows scientists to focus on creative and analytical tasks by reducing much of the tedious mechanics involved in computer work. The IA model best fits a distributed computing environment in which the IA hides the complexity of



*Metadata is at the core of the IA approach to improving data access.*

the environment but provides access to all of the computational power on the network.

The core concept of the IA approach is the creation and use of metadata, which is any information that helps scientists manage or interpret data. The IA design ensures that metadata reflects the content of data, aids data access, and is automatically generated when possible. A suite of integrated IA tools supports metadata creation, editing, browsing, and searching, in addition to file management.

## IA Tools and Prototypes

Wherever possible, the IA leverages existing software. Web tools and technologies such as search engines and hypermedia browsers are integral components of the IA. The cross-platform Java language is playing a key role in new development. Next, we will develop metadata tools for the end-user scientist, which are needed for scientific data management.

### *MetaMaker: Summarizing Calculations*

Efficient inspection and review of results from supercomputer calculations require filtering and summarization of the data. The MetaMaker tool summarizes calculational results by automatically generating metadata: text and image snapshots throughout a calculation. This metadata is presented to the user via a Web-browsable hypertext document.

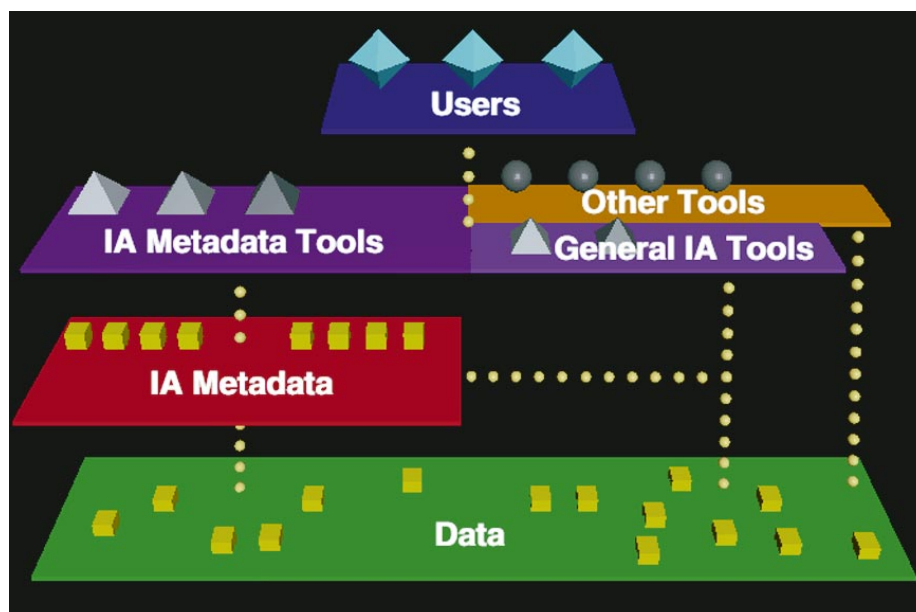
These thumbnail sketches with hypertext links to applications and data can be used to follow the calculation as it happens, to review final results, and to document the calculation for future reference.

### *DataMiner: Accessing Legacy Data*

Effective use of legacy data requires tools for browsing unfamiliar archives and generating relevant metadata. The prototype DataMiner tool supports interactive and automatic generation of metadata summaries. The graphical user interface guides users in specifying selected directories to traverse and summarize. Selected files are represented in one of three formats: content-based thumbnail sketches of all files in a directory; calculation summaries created by MetaMaker for selected files; and virtual directory reports, which represent relationships between files scattered across physical locations.

### *QuickNotes: Metadata Editing*

The QuickNotes tool provides a graphical user interface to support interactive browsing and editing of metadata for on-line files as well as physical and abstract objects. In addition to standard file system information such as owner, date, and size, QuickNotes displays content-based metadata including thumbnail graphics, pointers to related information, and user-defined abstracts and keywords. QuickNotes, written



IA tools support scientific data access and organization through the use of metadata.

## Related Projects

Project title	URL (Internet address)
<b>Intelligent Archive Project</b>	<a href="http://www.llnl.gov/ia/">http://www.llnl.gov/ia/</a>
<b>Livermore Computing Overview</b>	<a href="http://www.llnl.gov/liv_comp/">http://www.llnl.gov/liv_comp/</a>
<b>XDIR</b>	<a href="http://www.llnl.gov/ia/xdir.html">http://www.llnl.gov/ia/xdir.html</a>
<b>ASCI</b>	<a href="http://www.llnl.gov/liv_comp/asci/">http://www.llnl.gov/liv_comp/asci/</a>

in Java, uses a layout language to control which metadata fields are displayed and how they are presented. This allows metadata to be viewed in a way that can be customized for each type of user. The Hlist tool provides a graphical user interface for browsing hierarchical lists such as subject categories or a thesaurus.

### Metadata Manager

The Metadata Manager controls access to the metadata with a rela-

tional database management system (RDBMS). The other IA metadata tools rely on Metadata Manager to provide persistent storage and fast access to metadata. Searching can be accomplished directly through the RDBMS or with Web-compatible search engines operating on a static snapshot of metadata exported from the Metadata Manager.

### Searching

The search tool provides a convenient way for users to search and

browse through their data. A progressive technique allows the user to continually refine search parameters until the desired item or items have been located. This approach is also useful for browsing, since many related items are displayed together.

### XDIR: File Manager

With XDIR, a graphical FTP client and file manager, users can browse, find, and manipulate files in a heterogeneous network. It provides a graphical user interface for file transfer and for direct manipulation of local and remote directories on any FTP-addressable host. XDIR represents directories with tabular lists, long lists, tree structures, or icons and names. It can launch local and remote applications. It can also search local or remote directory structures for names that match a specified pattern.

## Collaborative Efforts

The core components of the IA are applicable to a variety of scientific disciplines. IA customers have included the Department of Energy (DOE) Defense Programs, the Accelerated Strategic Computing Initiative, and the Gas and Oil National Information Infrastructure.

*For more information about the IA, contact Jeff Long, IA Project Leader, 510-423-6421, [jwlong@llnl.gov](mailto:jwlong@llnl.gov).*